RICE UNIVERSITY

Testing the semantic control hypothesis for stroke aphasics with semantic deficits

by

Azli Hassan

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Arts

APPROVED, THESIS COMMITTEE

Randi Martin

Dr. Randi Martin, Committee Chair

Dr. Tatiana Schnur

Dr. David Lane

HOUSTON, TEXAS
April 2012

ABSTRACT

Testing the semantic control hypothesis for stroke aphasics with semantic deficits

by

Azli Hassan

Some studies of stroke patients with semantic deficits have found no effect of word frequency on semantic tasks, as well as inconsistent performance across items and tasks. A deficit in semantic control has been suggested as the source of the deficit - i.e., an inability to focus on semantic features appropriate to the task. In the present study, two stroke patients performed significantly better in single-distractor versions (low semantic control) than multiple-distractor versions of semantic tasks (high semantic control) of comprehension tasks, which appears consistent with the semantic control hypothesis. On the other hand, two aphasic patients showed substantially better performance for auditory than visual presentation of words in comprehension tasks – a finding that is not expected on the basis of semantic control. Experiment 1 evaluated whether performance on a multiple-distractor comprehension task could be predicted solely on the basis of performance on a single-distractor version using Luce's choice axiom. Single distractor performance significantly predicted performance and no convincing evidence was obtained for a role for semantic control. Experiment 2, which examined the modality effect, showed that for one of the patients, worse performance with auditory presentation was most likely due to rapid decay of phonological representations. For the other, worse performance was most likely due to a disruption to phonological representations of words

or to their connection to semantic representations. In all, the results suggest that word comprehension deficits in aphasia can result from a variety of sources and not all are due to semantic control deficits.

Acknowledgments

I would like to sincerely thank my advisor, Dr. Randi Martin, for her guidance and patience. And I would like to thank my family, friends (in particular, Anna, Amy, Aurora, Debshila, Jason, Yim Seong, and YingYing) , and Lilu (my cat) for all their love and encouragement.

Contents

List of Figures

List of Tables

Nomenclature

CCT	Camel and Cactus Test
EST	Environmental Sounds Task
LF	Left Frontal
LP	Left Parietal
LSA	Latent Semantic Analysis
LST	Left Superior Temporal
MCA	Middle Cerebral Artery

1. Introduction

The study of brain-damaged patients' behavioral deficits and lesion localizations allows us to attempt to determine the function of areas of the brain. As our brain is an interconnected network of regions, impairments in different areas may show similar effects at a global level but may have different underlying causes due to the functional role of the damaged areas, which can be uncovered through more specific testing. Jeffries and Lambon Ralph (2006) looked at semantic dementia patients and stroke aphasics with semantic deficits. Semantic dementia is a result of bilateral neurological deterioration, primarily in the temporal lobes, whereas stroke aphasia is a result of injury to language areas in the brain due to stroke. Although both groups were selected on the basis of demonstrating marked impairment on both verbal and nonverbal semantic tasks, Jefferies and Lambon Ralph found that their semantic dementia and stroke aphasic patients displayed different patterns on a number of factors related to semantic processing and concluded that the semantic dementia patients had a loss of semantic knowledge whereas the stroke aphasics had a disruption in access to semantic information. These differing patterns are discussed below.

1.1. Task Consistency

Semantic dementia patients displayed highly consistent performance on the same word items across all the tasks, whereas the stroke aphasics were only consistent across tasks requiring similar semantic judgments. One of the tasks they used, i.e. the Camel and Cactus Task (CCT, Bozeat et al, 2000), is a test of associative judgments between a probe (e.g. a picture of a Cactus) and four choices from which the participant must select the most related item: Cactus (target), Tree, Sunflower, and Rose. They used two versions of the CCT, one where the probe and items where presented as pictures, and another where they were presented as written words. In this task, the participant must first identify what aspect (e.g. where something is used or is usually found, what it can do) is relevant to matching the probe and target, while ignoring irrelevant aspects of the distractors and focusing on relevant aspects of the target that match the probe in order to make the correct associative judgment. Even though the two versions present the stimuli differently, it is assumed that the semantic control demands are similar. Another task, the Environmental Sounds Task (EST; Bozeat, Lambon Ralph, Patterson, Garrard, and Hodges, 2000) in which either sounds (e.g. a dog barking) or spoken words are presented and the subject must choose the matching target from nine categorically related distractors. In this task, all of the semantic features of the words and pictures are relevant for choosing the correct picture. Thus, it could be argued that the CCT and EST tasks differ in their semantic control requirements, since the CCT requires focusing on the relevant dimension whereas overall feature match is important for the EST task.

Jefferies and Lambon Ralph (2006) found significant correlations between performance on the picture and word versions of the CCT for both patient groups, and also between performance on the sound and spoken word versions of EST task. However, only the semantic dementia patients showed consistent performance between tasks with different semantic judgments; semantic dementia patients showed a significant correlation between performance between picture CCT and EST (Spoken Word to Pictures version), see Figure 1.

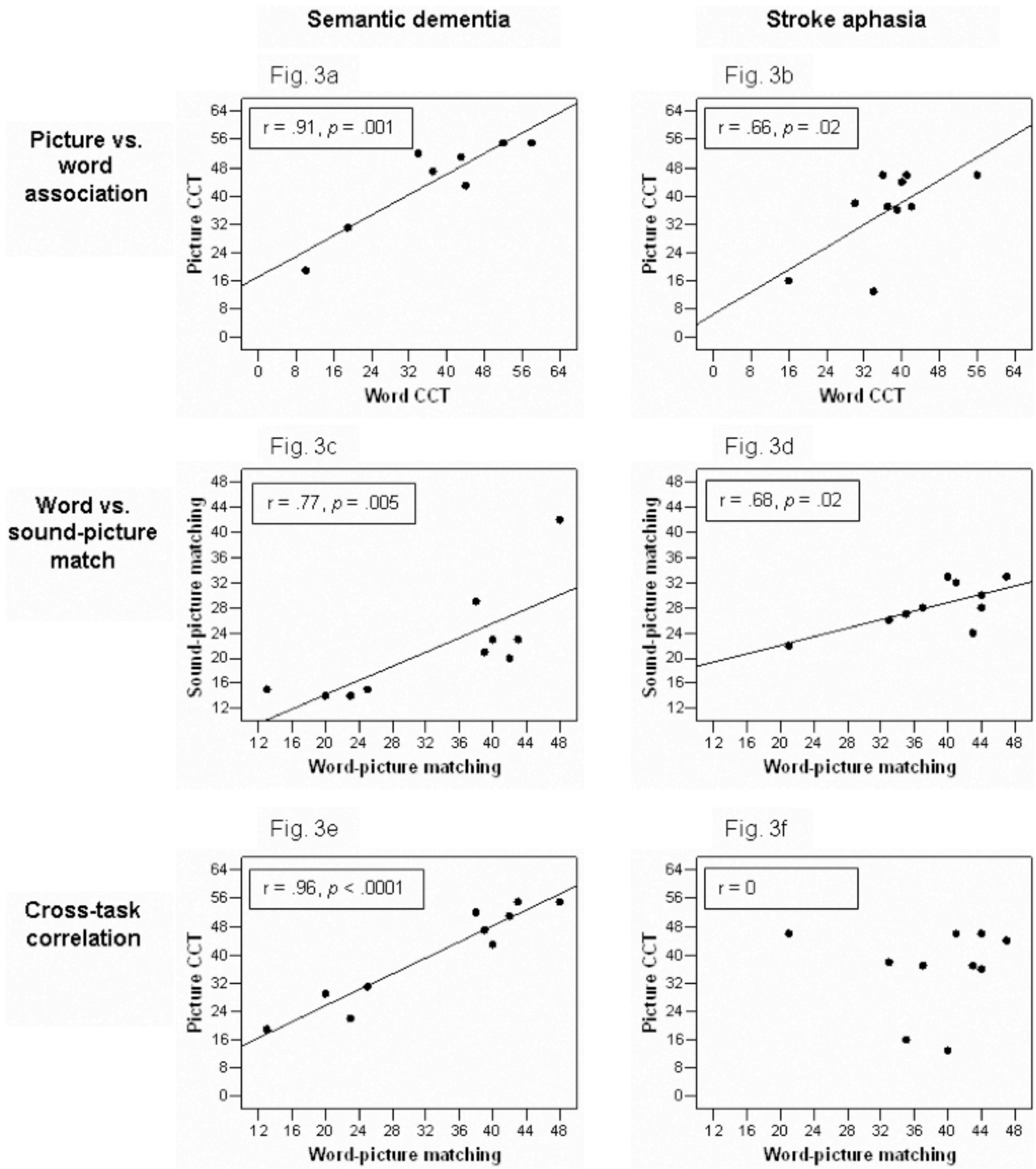


Figure 1. Correlations between task for semantic dementia patients and stroke aphasics (Jefferies, E., and Lambon-Ralph, M. 2006)

Based on these results, Jefferies and Lambon Ralph (2006) hypothesized that semantic dementia patients suffer from degraded semantic representations affecting all types of semantic knowledge from all modalities (e.g., visual knowledge, verbal knowledge) which results in high consistency in performance across comprehension tasks with different task demands. In contrast, they argued that the stroke aphasics suffer from impairment in semantic control, either in attending to relevant information or inhibiting irrelevant information, i.e. using semantic information to correctly perform the task. Since the semantic dementia patients have degraded amodal representations, the semantic cognition requirements do not matter since no representation is available regardless of the amount of cognitive control required. For the stroke aphasic patients, performance depends on the degree of cognitive control required.

1.2. Phonemic Cuing

In addition to the differing correlations across tasks for the two patient groups, a number of other findings were used to support their claim of a different basis for the semantic deficits in the two groups. For instance, Jefferies, Patterson, and Lambon-Ralph (2008) found that semantic dementia patients did not benefit as much as stroke aphasics from phonemic cueing in picture naming tasks (being given the first syllable of the name when the patient is unable to name a picture, e.g. /k/ for comb). They suggest that the increased activation of the phonological representation of the target name from the cue helps them overcome their impaired cognitive control, perhaps by allowing for easier resolution of conflict from competing responses. For the semantic dementia patients, the problem is not due to difficulty resolving conflict, but instead a failure to activate any names given a deficient semantic representation. Thus, cuing does not help.

1.3. Naming Errors

Additionally, Jefferies and Lambon Ralph (2006) argued that the type of errors made by each patient type provides some insight into the difference between the two groups. The semantic dementia patients made mostly coordinate-superordinate errors, e.g. when shown a picture of a zebra, the response given was 'horse' or 'animal', suggesting that their impairment stems from a gradual loss of semantic information. In addition to making coordinate-superordinate errors, the stroke aphasics made associative errors, e.g. responding with 'nut' to 'squirrel', suggesting that their impairment is with their semantic control, in that they are unable to inhibit a semantically related but inappropriate response.

1.4. Frequency Effect

Jefferies and colleagues have also found that, unlike the semantic dementia patients, their stroke aphasic patients showed no effect of word familiarity (Jefferies and Lambon Ralph, 2006) and no effect of word frequency on synonym judgments (Jefferies, Baker, Doran, Lambon Ralph, 2007). In addition, Jefferies et al. (2008) found a significant correlation between word frequency and naming accuracy for the semantic dementia patients but not for the stroke aphasic patients. They argue that the semantic dementia patients show word frequency effects because higher frequency words are more resilient to the effects of semantic degradation. They suggest that this is due to the progressive loss of semantic representations, which affects lower frequency items (which have weaker connections between features) before higher frequency connections (which have stronger connections). In the stroke patients, because of impaired semantic control, they are unable use the correct semantic information or associations, and had difficulty with rejecting distractors.

2. Problems with the semantic degradation vs. control hypothesis

2.1. Patterns of naming errors

While the findings reported by Jefferies and Lambon Ralph and colleagues (Jefferies & Lambon Ralph, 2006) provide evidence of interesting differences across the two patient groups, some other findings challenge the empirical basis of these claims. Budd et al. (2010) looked at aphasic patients with different areas of brain damage in relation to the left middle cerebral artery (MCA), i.e. Anterior MCA, Posterior MCA, as well as bilateral temporal atrophy in semantic dementia. They compared their picture naming performance to semantic dementia patients in order to assess the pattern of semantic coordinate errors. Budd et al. hypothesized that all patients would make such errors and that the type of errors would depend more on the location of brain damage rather than the etiology of the aphasia. Indeed, they found no significant difference between the groups in the rate of semantic coordinate errors and that the type of errors made depended more on location of brain damage than the nature of the disease. Budd et al. propose that amodal semantic degradation (as in Semantic Dementia) or degradation at the lexical-semantic representation (as in Stroke Aphasics) can lead to semantic coordinate errors in naming. However, Budd et al concede that both a semantic control explanation and a compensatory naming strategy (for those patients with relatively intact semantics) can explain the results of their own study. Additionally, unlike Jefferies and Lambon-Ralph (2006), Budd et al did not restrict their sample to patients with multimodal semantic deficits. They looked at stroke patients acutely after stroke whereas Jefferies and Lambon-Ralph looked at chronic aphasics.

2.2. Phonemic Cuing

For my first and second year project, we tested four stroke aphasic patients that presented with semantic deficits. Table 1 includes some background information on the patients we tested. These patients were tested on a word-picture matching task, e.g. being shown a picture of a comb, and then being asked “Is this a hairbrush?”. The item in the question could be the correct item, semantically related, phonologically related, or unrelated. Their inclusion in the study was based on impaired performance on semantically related items, while being relatively unimpaired in the other items (see Table 2).

Table 1

Stroke Patient background

	AC	**CA**	**DW**	**GB**
Age	49	42	54	84
Year of stroke	2000	2005	2000	2006
Speech	Non-Fluent	No Production	Fluent	No Production
Lesion	LF, LP & LST	LF, LP & LST	LF, incl. BA 44 & 45	LF, LP & LST

LF - Left Frontal, LP - Left Parietal, LST - Left Superior Temporal

Table 2

Word-Picture matching accuracy

	AC	CA	DW	GB
Semantically Related	74%	52%	83%	52%
Phonologically Related	100%	96%	96%	89%
Unrelated	100%	100%	100%	98%
Correct	100%	94%	100%	100%

The stroke patients with speech production , i.e. AC and DW, were tested on the Boston naming task (which involves naming a single picture) to determine if they showed any benefit from phonemic cuing. Patient AC was helped with phonemic cues while DW was not, see Table 3. Additionally, AC made at least 1 error of every type with the exception of a superordinate error, while DW made only circumlocution and omission errors, see Table 4.

Table 3

Boston naming results – Phonemic cueing

	AC	DW
Correct without phonemic cue	43	45
Correct with phonemic cue	9	0
Total Correct	52	45

Table 4

Boston naming results – Naming errors

Error Type	**AC**	**DW**
Coordinate	1	0
Superordinate	0	0
Associative	1	0
Phonological	1	0
Other	5	15
Total Correct	**52**	**45**

2.3. Semantic Control

Additionally, we compared performance in low and high semantic control conditions using the semantic tasks used in the Jefferies and Lambon Ralph (2006) study. The high semantic load conditions were identical (with a few changes to items due to cultural differences) to those used by Jefferies and Lambon Ralph (2006), whereas the low load conditions were the same tasks however instead of having multiple distractors, the low load conditions had a single distractor only. For instance, in the high load condition of the picture-word matching task, there were nine distractors, whereas in the low load condition there was only one distractor. We reasoned that more semantic control was required when the subject had to keep the target in mind while comparing it to each distractor. Comparing the data for multiple versus single distractor versions of the tasks, the stroke patients performed better when there was a single distractor (see Table 5; a complete summary of accuracy scores for each of the semantic tasks is in

Appendix A – Table A1). This is consistent with the Jefferies and Lambon Ralph's proposal that stroke aphasics have impairments in semantic control. That is, by decreasing the semantic control requirements, performance improved.

Table 5

Average accuracy across single and multiple distractor versions.

Condition	**AC**	**CA**	**DW**	**GB**
Single Distractor	93.0%	76%	93%	85%
Multiple Distractor	85.0%	48%	80%	60%
Difference	**8.0%**	**28%**	**14%**	**25%**

However, comparing accuracy between the low and high semantic control load conditions is not as straightforward as we initially thought since chance performance is different between the low and high load conditions. Additionally, given two trials with equal number of distractors in a task, we would also expect performance to differ based on how related the items are (i.e. more closely related items would be more difficult), so there is also the question of what level of performance would one expect given the relatedness of items to choose from. Assessing the difference between the high and low load conditions while taking into account the degree of relatedness between items and differing chance levels is one aim of the current study.

2.4. Sources of inconsistency across tasks

As discussed earlier, inconsistency in performance across tasks for aphasic patients but not for semantic dementia patients was one source of evidence used to argue for differing types of semantic deficits for the two patient groups. In the Brain and

Language Lab, we have observed a different type of inconsistency in stroke patients with aphasia, which is differing performance depending on the modality of the word input (i.e., for auditory vs. visual word input).

Specifically, two patients (CA and GB) performed significantly better ($p<.001$) on a word-picture matching task (seeing or hearing a word, then selecting the correct picture from nine semantically related pictures) when the word was presented as a written word versus a spoken word (see Table 6). Such a difference would not be expected for a semantic store deficit, but would not be expected for a semantic control deficit either. Similar kinds of semantic control would presumably be involved in mapping a spoken word and a written word to a corresponding object representation. Thus, it seems that other factors should be invoked to explain these differences across modalities.

Table 6

Overall accuracy on word-picture matching tasks

Task	**AC**	**CA**	**DW**	**GB**
Spoken word-picture matching - 1 Distractor	98%	81%	94%	78%
Spoken word-picture matching - 9 Distractors	92%	**38%**	95%	**50%**
Written word-picture matching - 1 Distractor	92%	94%	95%	94%
Written word-picture matching - 9 Distractors	94%	**84%**	94%	**95%**

One possibility is that for CA and GB that the fault lies in processing input phonology - i.e., that they are not processing the phonological information from the spoken word properly and thus have difficulty recognizing the words. As can be seen in

Table 6, their deficit with spoken input was particularly acute when they had to select from multiple distractors, which require maintaining the phonological information for a longer time period. A deficit in processing phonological input does not seem to be the complete explanation for CA and GB, however. All four patients performed well on unrelated trials and on phonologically related trials on single word-single picture matching tasks, e.g. presenting a picture of a cat and then asking if it is a picture of a hat (see Table 7). In fact, the patients that we tested were selected on the basis of good performance in the phonologically similar condition and poor performance in the semantically related condition (presenting a picture of a comb and then asking if it is a picture of a hairbrush), i.e. patients suspected of having a semantic deficit. However, both CA and GB show some impairment on consonant discrimination tasks, e.g. “pa-ba” (see Table 8), where each word in a pair differed on a single phonological feature (manner or articulation, place of articulation, or voicing). In an auditory lexical decision task, both CA and GB performed worse than AC and DW (See Table 9). Non-words (with a few exceptions) differed from real words on a single phonological feature (e.g. “bickle” derived from “pickle”). In fact, CA performed close to chance on the auditory lexical decision task. Her better performance on the phonologically related trials on picture-word matching task could be attributed to the fact that the phonologically related distractors in that task differed by more than a single phonological feature from the target word.

Table 7

Single word-single picture matching task accuracy scores

Condition	AC (Auditory)	CA (Auditory)	CA (Visual)	DW (Auditory)	GB (Auditory)	GB (Visual)
Semantically Related	74%	52%	50%	83%	52%	91%
Phonologically Similar	100%	96%	100%	96%	89%	100%
Unrelated	100%	100%	100%	100%	98%	100%
Correct	100%	94%	98%	100%	100%	100%

Table 8

Consonant discrimination accuracy scores

	AC	CA	DW	GB
Consonant-Vowel	96%	93%	93%	78%
Vowel-Consonant	93%	78%	91%	80%
Total	94%	85%	92%	79%

Table 9

Lexical decision accuracy scores

Modality	AC	CA	DW	GB
Auditory	93% (112/120)	56% (67/120)	89% (107/120)	77% (92/120)
Written	97% (116/120)	93% (111/120)	95% (114/120)	92% 110/120

Another possible contributing factor to CA and GB's poor performance on the semantically related trials may be that their imprecise phonological representations decay rapidly before a complete semantic representation is retrieved. For unrelated and phonologically related trials, this partial semantic activation is sufficient for choosing the correct picture. Alternatively, it is possible that even when the phonological-lexical representation does receive enough activation, the problem instead lies with retrieving the correct semantic representation (see Figure 2). Damage to the semantic representation itself is unlikely since they are performing well when the probe is presented as written stimuli. It is possible that connections between lexical-phonological and semantic representations are damaged such that, regardless of the activation at the phonological-lexical level, access to the semantic representation is impaired.

In order to determine whether rapid decay of the phonological representation is the source of the poor performance on the semantically related trials, we plan to manipulate the availability of the spoken and written words. If there is a problem in accessing semantic representations from lexical phonological representations, then increasing the availability of the spoken word should not help to increase performance on the spoken word version of the task. If rapid decay at the phonological level is the source, then increasing availability of the spoken word should improve performance. By manipulating the availability of the probe item, i.e. always available (continuous) vs. short presentation (limited), we hope to be able to tease apart where the deficit lies.

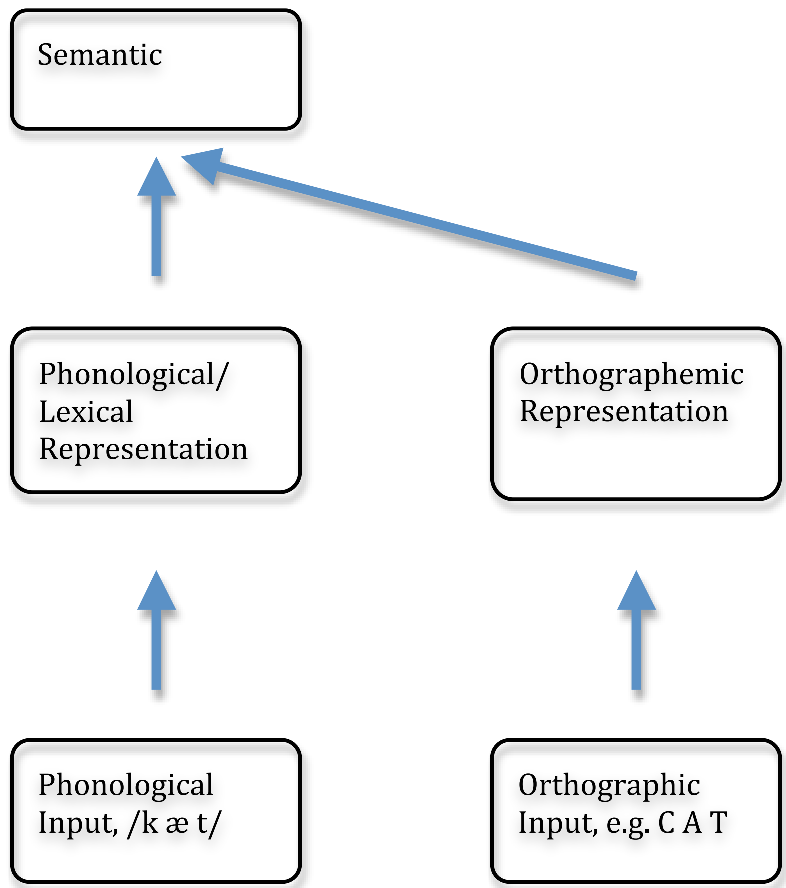


Figure 2. Representation of separate lexical representation for phonological and orthographemic inputs.

The current study serves to determine if the deficit in our two patients lies with maintaining an active representation of the probe or if the problem lies more with the retrieval from the semantic store. Additionally, it would also serve to determine if our patients were actually performing worse in the high executive function load condition versus the low load condition. We also hope to determine what factors make it harder to select among the potential answers, i.e. whether it is relatedness or similarity and how varying with the number of distractors interacts with relatedness/similarity. Perhaps having a single highly similar/related distractor may be as difficult or even more so than having multiple dissimilar/unrelated distractors.

3. The role of control in picture-word matching tasks

As discussed earlier, Jefferies and Lambon Ralph (2006) hypothesized that the lack of correlation across tasks for aphasic patients was due to the differing control requirements of the different tasks. However, in that study they did not specifically manipulate control requirements. One might hypothesize that selecting a target item among distractors requires executive function in order to select among all the items in semantic memory that are activated by the probe. If multiple semantically related representations are activated by the probe, then the more related items there are to select from, the more executive function ("semantic control") is required in order to be able to correctly select the target item. Previous research by Jefferies and Lambon Ralph (2006) found significant correlations between performance on semantic tasks and executive function in their stroke aphasic patients. The semantic dementia patients could not perform most of the executive function tasks (with the exception of the Ravens, which is typically thought to be an IQ test rather than an executive function test), and thus it was not possible to assess the correlation for these subjects but the semantic control hypothesis would predict no correlation between semantic task performance and executive functioning for the semantic dementia patients.

As mentioned earlier, the Brain and Language Lab looked at performance in stroke aphasic patients and found that overall, patients scored better when semantic control demands were lower, that is, when a single distractor was used, instead of multiple distractors. This finding is consistent with notion that executive function plays a role in semantic tasks.

However, in order to compare performance between versions of semantic tasks that include a single distractor versus multiple distractors, it is necessary to take into account the likelihood that a distractor picture will be judged to be more similar to the probe than the target picture on the relevant dimension (or dimensions) when there is a single distractor vs. multiple distractors. If the patients were performing at chance, chance performance would be 50% for tasks with a single distractor but only 10% for nine distractors. As shown in Table 5, the patients were clearly not performing at chance, as the lowest level of performance was 76.4% with a single distractor. However, let us assume that on each trial, subjects choose the picture that is most similar to the probe and that each distractor has some degree of similarity to the probe. If the similarity of one of the distractors is equal to or greater than that for the target, then the patient is likely to make an error. Evidence supporting this idea is what was found by Crutch and Warrington (2005). Crutch and Warrington found that performance for a stroke patient, described as having a refractory access deficit, decreased as semantic distance between items decreased in a word picture-matching task, i.e. the more closely related the items were, the more errors were made.

With multiple distractors, it is more likely that at least one picture distractor will have a similarity greater than the target than would be the case with a single distractor. In order to compute these likelihoods for multiple distractors, we need to know the similarity of the individual distractors. However, not all the distractors were presented in the single distractor conditions; thus, we do not have an estimate of similarity for each distractor for each patient. What is needed is some way to estimate the likelihood of a distractor being selected based on how related the distractor is to the target, such that the subject's performance can be compared against what would be predicted by similarity. Then a comparison can be made as to how actual performance compares to this estimate. If performance with multiple distractors is worse than this estimate, then the evidence for a contribution of semantic control would be stronger.

To determine if the semantic relatedness of the set of distractors in the multiple distractor conditions is plausibly related to patient performance, as a first pass, we assessed relatedness using Latent Semantic Analysis (LSA; Landauer and Dumais, 1997; Landauer, 1998). LSA is an objective measure of the semantic closeness of words in a semantic space. LSA can provide an objective, text-based measure of semantic relatedness. The basis for LSA is that a semantic space is created from a corpus of words, which accounts for the how words appear in different contexts and co-occurrence with other words in the same context. The co-occurrence values are used to determine a semantic space of specified dimensionality, essentially using a factor analysis of the co-occurrence values. In this way, a vector is created for each word in the semantic space and the cosine of the angle between the vectors is used as the measure of relatedness between two words. These LSA values have been successfully used to predict normal subjects' performance on a variety tasks involving semantic representations (Landauer, 1998).

To obtain a metric to relate to the probability of choosing the correct picture relative to set of distractors, we used Luce's Choice Axiom (see formula below), where the LSA values are used as weights representing the similarity of the distractors to the target. The weight for the target was set to 1.0. This provides an intuitively reasonable approximation of response selection. An example calculation is shown in Table 10, contrasting the probabilities between a set of semantically related and unrelated words.

$$P(i) = \frac{w_i}{\sum_j w_j}$$

where:

P(i): probability of selecting item i from a pool of j items

w: weight (a measure of some typically salient property) of a particular item

Table 10

Probabilities of choosing target picture of cat calculated using Luce's Choice Axiom and LSA values for a set of semantically related and unrelated words

(a) Related

Cat	1.00
Mouse	0.72
Squirrel	0.34
Cow	0.24
Frog	0.15
Rabbit	0.38
Sheep	0.00
Pig	0.33
Dog	0.36
Horse	0.06
Σ(1/LSA)	**0.28**

(b) Unrelated

Cat	1.00
Car	0.06
Truck	0.08
Motorcycle	0.03
Carriage	0.06
Airplane	0.01
Train	0.07
Bus	0.16
Bicycle	0.08
Sled	0.07
Σ(1/LSA)	**0.62**

These values could not be used to calculate the absolute values for choices for the patients since, for them, it is possible that some of the distractors had weights higher than or equal to those of the targets. Thus, we simply calculated these probabilities with Luce's Choice Axiom for each semantic category in the word-picture matching task (spoken and written version) and correlated them with the accuracy for that category for each stroke patient. The only significant correlation was for CA in the spoken version of the word-picture matching task, $r(6) = .78$, $p = .02$. It is should be noted that CA was the worst performer in the spoken version (38% correct) and showed a greater range of performance across categories. The lack of correlation for the other patients may have been due to a restriction of range. Alternatively, it is possible that the similarities from LSA do not accurately reflect similarity for the patients due to damage to specific semantic representations that alters the rank ordering of similarities from what is observed for normal subjects.

4. The role of frequency in semantic tasks for stroke patients

Unlike Jefferies and Lambon-Ralph we did find frequency effects with some of our patients. Specifically, DW and GB showed significant (or a trend towards significant) frequency effects (Table A2 and A3), whereas AC and CA did not. Hoffman, Rogers, and Lambon Ralph (2011) calculated a semantic diversity measure, based on context vectors rather than co-occurrence vectors using LSA, that accounted for a statistically significant amount of variation in item performance in stroke aphasic patients. Their semantic diversity measure was correlated with word frequency, i.e. high frequency words tended to be more semantically diverse. It was also correlated with imageability in that less imageable (more abstract) words are more semantically diverse. What is interesting is that, without semantic diversity measure being included in a multiple regression model that included frequency and imageability, frequency was not a significant predictor. When semantic diversity was included, frequency was a significant predictor albeit the weakest one while semantic diversity was the strongest predictor. They conclude that the frequency in stroke aphasics is masked by this semantic diversity effect.

As discussed earlier, the main focus of this study is to test the semantic control hypothesis of Jefferies and Lambon Ralph (2006) for semantic deficits in stroke aphasics. This will be accomplished by testing pair-wise choice comparisons of all targets and distractors to determine if the probability of correct choices in the multiple distractor condition can be completely determined by pair-wise choices. If so, this would be evidence against the semantic control hypothesis. In investigating this issue, we will focus on patients CA and DW. Although both CA and GB showed large differences between performance with single vs. multiple distractors, these two patients showed the large discrepancy between performance with auditory vs. written words. For GB, large performance decrements with multiple distractors occurred only with auditory presentation. A rapid decay of phonological information may be the source of their deficit with spoken words. Such a deficit would be expected to cause difficulties when having to hold a probe word in mind while searching through multiple distractors. For CA and DW, discrepancies were observed between the single and multiple distractor conditions for the CCT for both picture and (written) word versions. (GB did not show a discrepancy for the picture version of the CCT.) Thus, CA and DW will be tested on pair-wise choices for the targets and distractors using the picture version of the CCT task.

Also, we will investigate the source of the discrepancy between performance for the written and spoken modalities for CA and GB to determine if a rapid loss of phonological information or impaired access from lexical to semantic representations is the source of their poor performance for auditory presentation.

5. Experiments 1 and 2

5.1. Method and Subjects

5.1.1. Subjects.

Patients CA and DW were tested in Experiment 1, and patients CA and GB were tested in Experiment 2.

5.1.2. Materials, Design, and Procedure.

In Experiment 1, target pictures and distractors from the picture version of the CCT were used. CA and DW were retested on the multiple distractor version of the task since it has been a long time since they were previously tested. Additionally, they were tested three times on all possible pairs of the target and distractor for each probe. On each trial, the probe appeared above the target and distractor. The target and distractor were arranged horizontally. As there are 64 trials on the CCT, presenting each target–distractor pair results in 192 trials. Three separate lists were created, each presenting the 192 trials once. The trials alternated between a living and non-living category, see Table 11. Each of the 64 items was run through once with a distractor, before repeating the same sequence a second and third time with the remaining distractors. Each distractor appeared in the first, second, and third cycle across each list, see Table 12. For each list, the side on which the target appeared was randomized such that each side was equally likely to contain the target.

Table 11.

Order of categories for each list

List 1	**List 2**	**List 3**
Birds	Small Household	Fruits
Large Household	Fruits	Vehicles
Domestic Animals	Large Household	Foreign Animals
Small Household	Birds	Tools
Foreign Animals	Vehicles	Domestic Animals
Tools	Domestic Animals	Small Household
Fruits	Tools	Birds
Vehicles	Foreign Animals	Large Household

Table 12.

Presentation order for distractors

List 1	**List 2**	**List 3**
Distractor 1	Distractor 3	Distractor 2
Distractor 2	Distractor 1	Distractor 3
Distractor 3	Distractor 2	Distractor 1

In Experiment 2, both CA and GB were tested on modified versions of the multiple distractor word picture matching task, once with auditory presentation and once with visual presentation. For the modified spoken version, the subjects were able to press a button to hear the word as many times as they wish until they made a response. For the modified written version, the probe was displayed onscreen for only 750 ms. For each version, there were 64 trials.

5.2. Results.

5.2.1. Experiment 1 – Pair-wise Comparisons for CCT

Overall accuracy across all the pair-wise trials was relatively high (CA: 84%; DW: 93%). Figure 3 shows the average accuracy for each list, while Figure 4 shows the average accuracy for each presentation order of distractors (e.g. Presentation 1 is the average of the first target-distractor pair across all three lists). CA showed some increase in accuracy across each list as well as an increase on the last presentation of a target-distractor pair, while DW was consistent across lists and presentations. For the multiple distractor version of the CCT, CA got 59% of the trials correct, while DW got 77%. Thus, it is clear that both patients did substantially better on the pair-wise trials than on the multiple-distractor trials.

(a) CA (b) DW

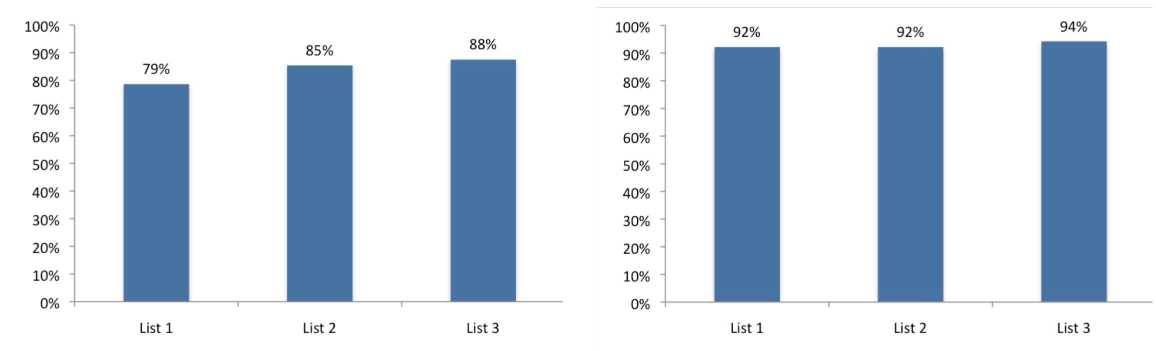


Figure 3. Average accuracy for each pair-wise judgment list

(a) CA (b) DW

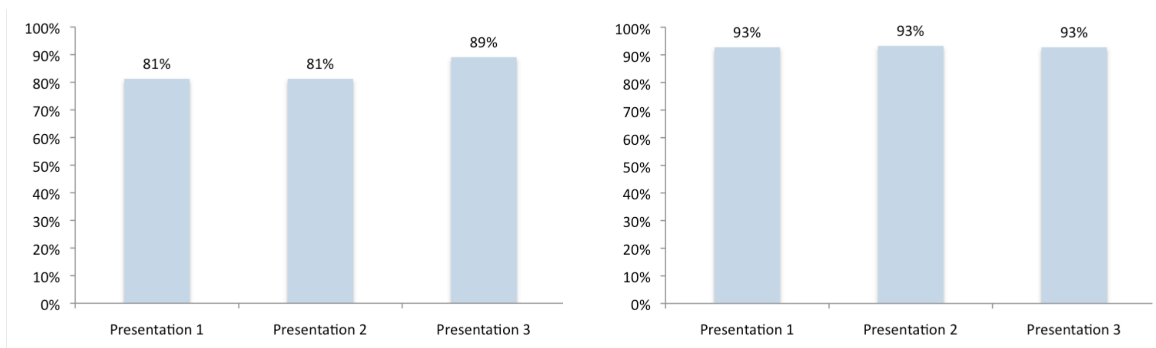


Figure 4. Average accuracy for each distractor presentation order

Average accuracy was calculated for each trial using the new administration of the multiple distractor CCT for this study and the previous administration resulting in possible values of 0, 0.5, and 1. For both subjects, the correlation between the old and new CCT multiple distractor answers was significant (CA: $r(62) = .32$, $p = .01$; DW: $r(62) = .48$, $p < .001$). In order to determine if the probabilities of correct choices from the pair-wise decisions predicted multiple distractor performance, a linear regression was carried out with the Luce choice probability derived from the pair-wise accuracies (pair-wise), log transformed word frequency (logfreq), and their interaction. However, as the interaction did not come close to reaching significance (p > .22 for both patients), the interaction was eliminated from the regression equation. In computing the Luce choice probability, the similarity value for the target to the correct picture choice was assumed to be one whereas the similarity to the distractors was assumed to be 1- proportion correct for each distractor. The results are shown in Table 13 for both patients, the overall regression was significant (CA: R^2 (61) = .45, $F(2)$ =24.70, $p < .001$.; DW: R^2 (61) = .22, $F(2)$ = 8.31, $p = .001$). For both CA and DW, pair-wise was the only significant predictor (CA: $t(61) = 6.69$, $p < .001$; DW: $t(61) = 3.00$, $p < .001$).

Table 13.

Regression coefficients and statistics

(a) CA

Term	**B**	**Std. Error**	***t***	***p***
Constant	-0.46	0.24	-1.93	0.06
Pair-wise	1.01	0.15	6.69	<.001
LogFreq	0.08	0.06	1.37	0.18

(b) DW

Term	B	Std. Error	*t*	*p*
Constant	-0.35	0.33	-1.07	0.29
Pair-wise	0.72	0.24	3.00	<.01
Log(Freq)	0.12	0.07	1.60	0.12

The same analysis was run by assuming to the strength of the target to the probe to be equal to the average accuracy across all the pair-wise judgment trials (with possible values ranging from 0 to 1). The overall picture was the same in that the models with pair-wise and logfreq were significantly better than baseline with pair-wise being the only significant predictor for both subjects. The similarity of the results is not surprising, given that the two calculations of Luce choice probability were highly correlated with each other ($r(62) = .997$, $p < .001$).

The results showed that for both patients, pair-wise accuracy predicted multiple-choice performance, unlike what was found for DW when using the LSA values for prediction. Thus, at least in the case of DW, similarities derived from pair-wise choices did a better job in predicting performance than similarities derived from general text-base co-occurrence (which is the basis of the LSA values). However, a correlation between pair-wise choices and multiple-choice performance would be expected, both if performance on pair-wise choices completely determined multiple distractor performance or if semantic control also played a role. That is, even if semantic control <u>is</u> important, one would still expect the relative similarity of the distractors to influence the difficulty of making choices. If semantic control plays an additional role, one would predict that performance on the multiple-choice distractors to be lower than that predicted by the pair-wise choices. Table 14 shows the accuracy of multiple choice for different values of the pair-wise choices. To address the issue of the degree of over-prediction of performance based on pair-wise choices, I examined the regression equation predicting multiple-choice performance from pair-wise choices. If performance is lower than predicted, one would expect that either the intercept would be below zero or the slope would be less than one. Figure 5 shows the data with the regression line derived from the data as well as the predicted performance for CA and DW if pair-wise choices perfectly predicted multiple choices according to the Luce choice axiom. A bubble chart was used to show how many data points lie at each point (i.e. the area of the bubble is proportional to the number of data points). For both CA and DW, the majority of data points are at the point where the average accuracy and pair-wise predictor are both equal to 1 (CA: 25/64; DW: 32/64).

Table 14.

Count of data points for values of pair-wise and average accuracy

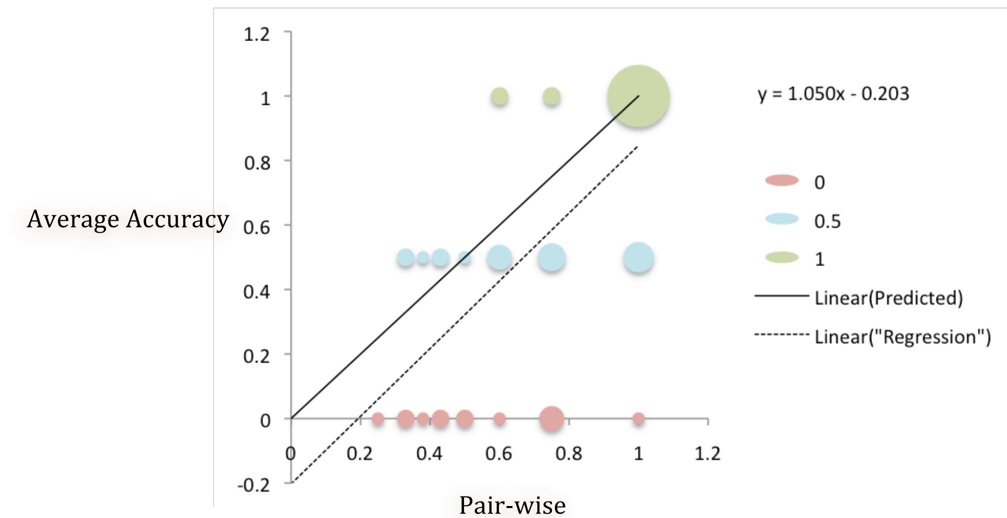
(a) CA

Pair-wise	**Average Accuracy**		
	0	*0.5*	*1*
0.25	1		
0.33	2	2	
0.38	1	1	
0.43	2	2	
0.5	2	1	
0.6	1	4	2
0.75	4	5	2
1	1	6	25

(b) DW

Pair-wise	**Average Accuracy**		
	0	*0.5*	*1*
0.25	1		
0.27	1		
0.33		1	
0.43		1	
0.5	1		
0.6		1	
0.75	3	1	5
1	5	12	32

(a) CA



(b) DW

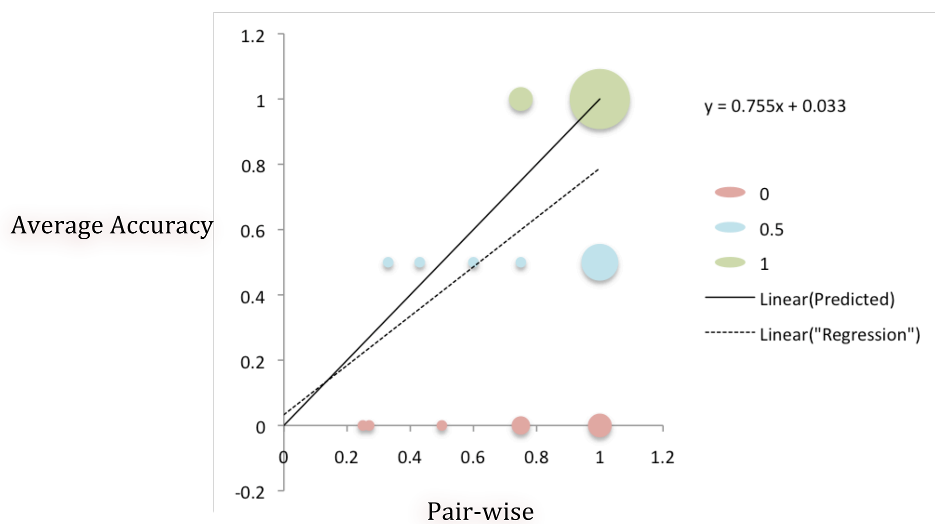


Figure 5. Bubble chart of average accuracy and pair-wise predictor with actual and predicted regression lines.

For both CA, the slope was very close to 1.0 (b = 1.05, $t(62) = .33$, $p = .37$). The intercept was below zero (a = -.203), and showed a trend towards significance ($t(62) = 1.64$, $p = .053$). For DW, the slope was less than 1.0 (b = .755) but not significantly so ($t(62) = 1.03$, $p = .15$) and the intercept was not significantly different than 0 (a= -.03, $t(62) = .15$, $p = .44$). All regression slope and intercept t-tests were one-tailed. Thus, for both patients, no strong evidence was obtained that multiple distractor performance was lower than that predicted by pair-wise distractor performance.

5.2.2. Experiment 2 – Standard and Modified Word-Picture Matching.

In the written version of the Word-Picture Matching task, CA showed a significant drop in accuracy, 81% vs. 64%, when the probe was only available for 750 ms ($t(62) = 3.01$, $p < .01$). GB showed no significant difference between the standard and modified version (92% vs. 90%), see Figure 6.

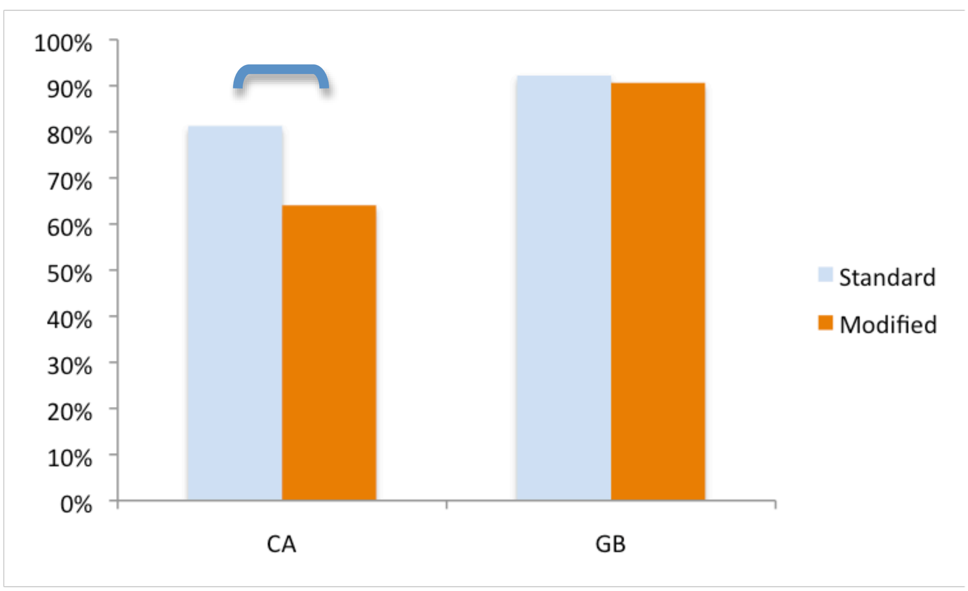


Figure 6. Accuracy for written version of Word-Picture Matching task

In the spoken version of the Word-Picture Matching task, GB showed a increase in accuracy (42% vs. 56%) when the probe was available for repeat listening that showed a trend towards significance ($t(62) = 1.83$, $p = .07$). CA showed no significant difference between the standard and modified version (52% vs. 53%, see Figure 7).

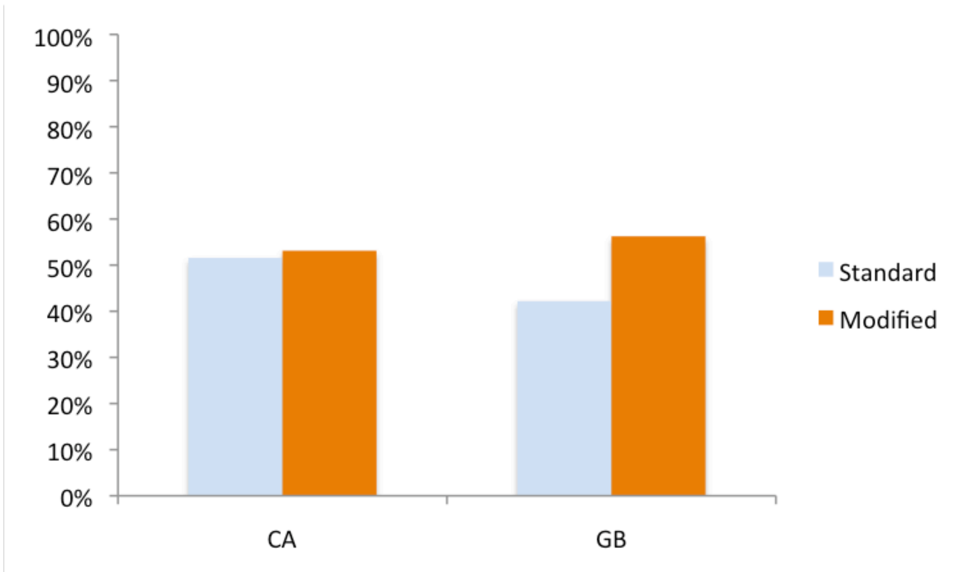


Figure 7. Accuracy for spoken version of Word-Picture Matching task

6. Discussion

In Experiment 1, we tested the hypothesis that relatedness of the target to distractors in the CCT predicts performance between the single and multiple distractor versions rather than an impairment in semantic control. As expected, for both aphasic subjects (CA and DW), their pair-wise judgments with single distractors were a significant predictor of performance in the multiple distractor version . Critically, there was no strong evidence that performance with multiple distractors was worse than would be expected based on pair-wise performance. This strongly suggests that for performance in the multiple distractor version of the CCT, it is the relatedness of items that affects performance and not the number of items, which is inconsistent with the semantic control hypothesis. The semantic control hypothesis would predict that an increase in the number of distractors would make it harder to select the correct response, whereas our results suggest that our patients are using the own internal relatedness weights to select an answer when there is more than one distractor. Recall that when the pair-wise predictor is 1, it means that the subject is has correctly answered each pair-wise judgment across all three lists. Given that the majority of correct responses occur when the pair-wise predictor is close to one, this suggests that the subjects are, for the most part, correctly using their own target-distractor comparisons when there are multiple distractors. It is possible that there is a failure in semantic control due to attending to the incorrect attribute or an inability to retrieve the correct representation in the pair-wise judgments, but a failure in semantic control should be exacerbated with an increase items that have to be considered for a response and this was not obviously the case in this study.

While word frequency was not a significant predictor of multiple choice performance for either CA or DW when pair-wise performance was included as a predictor, there was still some suggestion of a frequency effect for DW ($p = .12$), which is interesting given that she had shown significant effects of frequency on various tasks, including on the first administration of the multiple-choice word version of the CCT test that was employed in this experiment (see Table A2). The lack of significance of frequency might potentially be attributed to collinearity between the pair-wise choices and probe frequency. However, these correlations were quite small for both patients (CA:

$r(61) = .12$, $p = .37$; DW: $r(61) = .08$, $p = .52$). For CA, it may be that effects of word frequency are being masked by other effects, perhaps semantic diversity (Hoffman, Rogers, and Lambon Ralph, 2011). An avenue for future investigation would be to create test stimuli where relatedness, word frequency, and semantic diversity are better controlled for to see if word frequency would show up as significant predictor when the other two factors are controlled for.

One avenue of research that may follow this study would be to compare the semantic control requirements for different tasks in terms of the relative relatedness of target and distractors across tasks. For example, a comparison of LSA values for items in the word-picture matching task and the CCT could be performed. We expect that for the CCT relative to word-picture matching, the LSA values for relatedness amongst target and distrators in the CCT will be larger (i.e. more related/similar), and that the LSA values between probe and target will be smaller. In comparing the word-picture matching tasks with CCT, we expect to see that the difference in performance (i.e. worse performance in the CCT than the word-picture matching task) is due to target and distractor items being highly related to each other, while the target is less related to the probe. It is also true that the nature of the task is different. In the word-picture matching task, it is likely that all semantic features are used to correctly identify the target, whereas in the CCT, it is a single dimension that must be correctly identified in order to select the target. Focusing in on a single feature would seem to require greater semantic control.

In Experiment 2, we tested the hypothesis that poorer performance in the spoken version of the word-picture matching task is due to having to maintain the spoken word in memory as it is only heard once, whereas in the written version the word appears on the screen till an answer is given and therefore no maintenance is required. Thus, the modified versions of the task were designed to reverse the availability of the probe in both the spoken and written versions of the task. Somewhat surprisingly, the two patients showed different patterns of performance between the standard and modified versions of the word-picture matching task.

For patient CA, reducing the availability of the written probe results in a significant drop in accuracy. However, increasing the availability of the spoken probe did not improve her accuracy. In contrast, GB showed the opposite pattern to CA, in that reducing the availability of the written probe did not significantly affect his performance, while increasing the availability of the spoken probe resulted in an increase in accuracy that showed a trend towards significance.

Given that CA is able to perform relatively well when the probe is available till an answer is given, it is likely that the problem with the brief visual presentation is with holding onto the probe in memory or perhaps with recognizing the word in the shorter amount of time given, instead of it being a problem at the level of the orthographemic representation. GB showed no detriment with brief visual performance and performed well regardless of the availability of the probe, and as such shows no impairment to recognizing the written word nor in maintaining the probe in memory.

CA showed no benefit with increased availability of the spoken probe word. Given her poor performance in both versions of the spoken WPM, the results suggest that CA has damaged connections between phonological/lexical representations and semantic representations since it is unlikely that her difficulties are due to phonological processing disruption, given her good performance with phonologically related distractors in picture-word matching. It is unlikely that rapid decay of phonological information can explain her poor performance with auditory presentation since her performance did not improve when the probe was available for repeated listening. GB showed a benefit with increased availability of the probe, albeit not a statistically significant benefit, suggesting that at least part of his poor performance in the standard version of the spoken WPM is due to problems with holding onto the spoken word in memory.

In summary, the current study suggests that poorer performance on the multiple distractor condition relative to the pair-wise condition was not due to differing semantic control requirements, but could be accounted for by the relatedness of distractors to the target and the greater likelihood of having a highly related distractor in the multiple choice condition. As such, the current study does not provide evidence to support the semantic control hypothesis as put forth by Jefferies and Lambon Ralph (2006). Additionally, performance in tasks with similar semantic control requirements can differ based on the input modality, and in this study that difference seems to arise from individual impairments to two separate representation stores, i.e. phonological/lexical representations and orthographemic representations. Thus, semantic deficits in aphasia can derive from different sources – not all of which are related to semantic control.

References

Bozeat, S., Lambon Ralph, M.A., Patterson, P., Garrad, P., and Hodges, J.R. (2000) Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, 38 (9), 1207-1215

Crutch. S.J., and Warrington, E.K. (2005). Gradients of semantic relatedness and their contrasting explanations in refractory access and storage semantic impairments. *Cognitive Neuropsychology*, 27 (7), 851-876.

Hoffman, P., Rogers, T.T., and Lambon Ralph, M.A. (2011) Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience,* 23 (9), 2432-46

Jefferies. E, Baker. S.S., Doran. M, and Lambon Ralph, M.A. (2007). Refractory effects in stroke aphasia: A consequence of poor semantic control. *Neuropsychologia*, 45, 1065-1079.

Jefferies, E., and Lambon Ralph, M.A. (2006) Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain : A Journal of Neurology*, 129 (8), 2132-47

Jefferies. E, Patterson, K, and Lambon Ralph. M.A. (2008). Deficits of knowledge versus executive control in semantic cognition: Insights from cued naming. *Neuropsychologia*, 46, 649-658.

Landauer, T.K. (1998) Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science*, 7 (5), 161-164

Landauer, T.K., and Dumais, S.T. (1997) A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104 (2), 211-240

Rapp, B., and Caramazza, A. (1993) On the distinction between deficits of access and deficits of storage: A question of theory. *Cognitive Neuropsychology*, 10, 2, 113 - 141.

Appendix A - Semantic Tasks Results

Table A1

Overall Accuracy Scores

Task	AC	CA	DW	GB
Camel and Cactus Test (Pic CCT) - 1 Distractor	83%	66%	**85%**	73%
Camel and Cactus Test (Pic CCT) - 3 Distractors	81%	61%	**67%**	75%
Camel and Cactus Test (Word CCT) - 1 Distractor	88%	**69%**	**92%**	**81%**
Camel and Cactus Test (Word CCT) - 3 Distractors	78%	**59%**	**83%**	**59%**
Environmental Sounds Battery (sounds to pictures - 1 Distractor)	96%	**83%**	**90%**	**85%**
Environmental Sounds Battery (sounds to pictures - 9 Distractors)	79%	**35%**	**71%**	**63%**
Environmental Sounds Battery (sounds to written words - 1 Distractor)	**100%**	**83%**	**92%**	**88%**
Environmental Sounds Battery (sounds to written words - 9 Distractors)	**79%**	**42%**	**67%**	**50%**
Environmental Sounds Battery (spoken words to pictures - 1 Distractor)	100%	**92%**	**100%**	**85%**
Environmental Sounds Battery (spoken words to pictures - 9 Distractors)	96%	**46%**	**92%**	**46%**
Environmental Sounds Battery (written words to pictures - 1 Distractor)	100%	**92%**	100%	85%
Environmental Sounds Battery (written words to pictures - 9 Distractors)	96%	**62%**	90%	88%
Spoken word-picture matching - 1 Distractor	98%	**81%**	94%	**78%**
Spoken word-picture matching - 9 Distractors	92%	**38%**	95%	**50%**
Written word-picture matching - 1 Distractor	92%	94%	95%	94%
Written word-picture matching - 9 Distractors	94%	84%	94%	95%

Bold font *denotes a significant difference between the single and multiple distractor conditions*

Table A2

Single Distractor tasks - Accuracy scores for high and low frequency words and p-values for the difference

Task (Single Distractor)	**Frequency**	**AC**	**CA**	**DW**	**GB**
CCT - Pic	Low	75%	65%	85%	60%
	High	90%	70%	85%	75%
	p	*0.44*	*0.74*	*1*	*0.32*
CCT - Word	Low	80%	65%	85%	85%
	High	95%	70%	95%	90%
	p	*0.16*	*0.74*	*0.30*	*0.64*
Env Snd (Sounds-Pic)	Low	95%	85%	73%	80%
	High	100%	80%	93%	80%
	p	*0.38*	*0.69*	*0.18*	*1*
Env Snd (Sounds-Written)	Low	70%	80%	87%	93%
	High	80%	64%	100%	93%
	p	*0.63*	*0.32*	*0.15*	*1*
Env Snd (Spoken-Pic)	Low	100%	90%	100%	80%
	High	100%	80%	93%	86%
	p	*1*	*0.56*	*0.31*	*0.70*
Env Snd (Written-Pic)	Low	100%	93%	100%	80%
	High	100%	86%	100%	86%
	p	*1*	*0.51*	*1*	*0.697*
Spoken-Pic Match	Low	95%	75%	85%	80%
	High	100%	85%	100%	85%
	p	*0.32*	*0.44*	***0.08***	*0.69*
Written-Pic Match	Low	90%	90%	95%	90%
	High	90%	95%	95%	95%
	p	*1*	*0.56*	*1*	*0.56*

Bolded p-values show a trend towards significance, i.e. $p < .1$

Table A2

Multiple Distractor tasks - Accuracy scores for high and low frequency words and p-values for the difference

Task (Multiple Distractors)	Frequency	AC	CA	DW	GB
CCT - Pic	Low	80%	40%	55%	65%
	High	75%	50%	85%	70%
	p	*0.71*	*0.76*	***0.04***	*0.74*
CCT - Word	Low	65%	45%	70%	35%
	High	80%	70%	95%	65%
	p	*0.3*	*0.12*	***0.04***	***0.06***
Env Snd (Sounds-Pic)	Low	88%	30%	40%	73%
	High	73%	35%	80%	40%
	p	*0.38*	*0.74*	***0.03***	***0.07****
Env Snd (Sounds-Written)	Low	80%	30%	40%	40%
	High	80%	36%	73%	79%
	p	*1*	*0.74*	***0.07***	***0.04***
Env Snd (Spoken-Pic)	Low	90%	30%	100%	60%
	High	100%	50%	100%	40%
	p	*0.33*	*0.39*	*1*	*0.29*
Env Snd (Written-Pic)	Low	100%	47%	93%	87%
	High	100%	64%	93%	93%
	p	*1*	*0.36*	*1*	*0.6*
Spoken-Pic Match	Low	95%	45%	95%	45%
	High	100%	30%	95%	30%
	p	*0.33*	*0.34*	*1*	*0.34*
Written-Pic Match	Low	85%	70%	95%	85%
	High	95%	80%	100%	100%
	p	*0.3*	*0.48*	*0.32*	***0.08***

*Bolded p-values show a trend towards significance, i.e. $p < .1$, and shaded p-values are significant, i.e. $p < .05$ (*This result is in the opposite direction of the standard frequency effect)*